

Antonis Maronikolakis, PhD Student

[github](#)[email](#)[linkedin](#)[website](#)[scholar](#)

🏛 Education

MSc Speech and Language Processing - First Honours (1:1)

September 2018 - September 2019

University of Sheffield, Sheffield, UK

Thesis: Text Generation for News Headlines

BSc Computer Science - 8.4/10.0 (Upper 2:1)

October 2014 - June 2018

University of Piraeus, Athens, Greece

Thesis: Approximation and Heuristic Algorithms for Ridesharing

🔧 Technologies

Libraries: HuggingFace, Sklearn, Pandas, Keras, NLTK, PyTorch

Tools: LaTeX, VS Code, Bash, Git

🗣 Research Interests

hate speech, social nlp, model and data bias, dataset curation, data ethics

🎓 PhD Student (Start of 3rd Year)

- Supervised by *Hinrich Schütze*, *Ludwig-Maximilians-Universität München* January 2020 - current

- Currently working on hate speech analysis in a multi-disciplinary project, [AI4Dignity](#), to tackle hate speech against marginalized communities around the world, in an effort that involves the curation and analysis of hateful data.
- Leading side and student projects, publishing two workshop papers in the field of generated text analysis and multi-domain pre-training.
- Previously investigated multilinguality in BERT, examining the effect of tokenization compatibility on multilinguality (EMNLP 2021 Findings) and showing that BERT cannot align characters across languages (EMNLP 2021 “Workshop on Insights from Negative Results”).

📖 Publications

[Wine is Not v i n. – On the Compatibility of Tokenizations Across Languages](#). Antonis Maronikolakis, Philipp Dufter, and Hinrich Schütze. 2021. In Findings of the Association for Computational Linguistics: EMNLP 2021, Punta Cana, Dominican Republic. Association for Computational Linguistics.

[BERT Cannot Align Characters](#). Antonis Maronikolakis, Philipp Dufter, and Hinrich Schütze. 2021. In Proceedings of the Second Workshop on Insights from Negative Results in NLP, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[Artificial Intelligence, Extreme Speech, and the Challenges of Online Content Moderation](#). Sahana Udupa, Elonnai Hickok, Antonis Maronikolakis, Hinrich Schuetze, Laura Csuka, Axel Wisiosek, Leah Nann. 2021. AI4Dignity Project. EU Commission Policy Brief.

[Identifying Automatically Generated Headlines Using Transformers](#). Antonis Maronikolakis, Hinrich Schütze, and Mark Stevenson. 2021. In Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda. Association for Computational Linguistics.

[Multidomain Pretrained Language Models for Green NLP](#). Antonis Maronikolakis and Hinrich Schütze. 2021. In Proceedings of the Second Workshop on Domain Adaptation for NLP, Kyiv, Ukraine. Association for Computational Linguistics.

[Analyzing Political Parody in Social Media](#). Antonis Maronikolakis, Danae Sánchez Villegas, Daniel Preotiuc-Pietro, and Nikolaos Aletras. 2020. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

🔗 Professional Experience

NLP Researcher for [AI4Dignity](#)

- *Extreme Speech Analysis for Marginalized Communities*

November 2020 - December 2021

- Providing guidance from an NLP perspective in a multi-disciplinary team of NLP researchers, anthropologists and fact-checkers.
- Led data collection and annotation efforts culminating in a dataset of 20k examples and multiple label categories from a team of 13 annotators spanning 4 countries.
- Organized a two-day online Counterathon event where academics, NLP practitioners and fact-checkers sat down in teams to brainstorm ideas pertaining to extreme speech online.
- Gave two talks introducing our project as a keynote speaker at [MMHS](#) and opening as a speaker on our Counterathon event.

Speech Processing Practitioner for [VoiceBase](#) Research Lab

- *Analyzed Variational Autoencoders and Phrase Detection methods*

November 2018 - August 2019

- As part of Dr. Thomas Hain's VoiceBase research lab, I was assisting researchers by exploring a novel Variational Autoencoder structure, alongside work on Phrase Detection and Extraction.
- Analyzed and ran experiments on the Factorized Hierarchical Variational Autoencoder [repository](#) and on various Phrase Detection algorithms.
- Set up and documented experiment configurations for [ESPN](#) and [DeepSpeech](#).

AI Programmer/Writer for Google Summer of Code

- *Mentored under Dr. Peter Norvig, working on AI algorithms*

June - September 2017

- Worked on the [Python repository](#) of Dr. Norvig's book, *Artificial Intelligence: A Modern Approach*.
- Implemented, evaluated and wrote on Natural Language Processing and Machine Learning algorithms and concepts, explaining how they work and providing examples students can work with.
- Collaborated with Dr. Norvig to polish pseudocode from the book.
- After the end of the project, I was given administrator rights to the repository.

🏆 Awards

- Scholarship for academic excellence - 2016-2017, 2017-2018
- Kaggle Machine Learning Competitions (2018) - Silver and Bronze Medals
- Microsoft Imagine Cup Competition (2015 and 2016) - National Finalist
- National Computer Science Competition 2014 - Finalist